社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－DBS－141 (2)
2007－GN－ 62 (2)
2007／1／25

ユーザ属性と検索領域知識の関連文書提示による意思決定支援システム

大塚 英史† 西園 敏弘‡

†‡日本大学工学部 〒963-8642 福島県郡山市田村町徳定字中河原1番地
E-mail： †otsuka@cscw00.ce.nihon-u.ac.jp ‡nishizono@cs.ce.nihon-u.ac.jp

あらまし 問題解決のヒントとなる文書を提示することにより，意思決定プロセスの初期段階を支援するシステムを提案する．この段階では，選択肢が広範囲に及ぶため，問題解決に必要な情報を検索するために，ユーザの意図を的確なキーワードで示すことが難しい．本システムは，趣味や仕事上の役職などのユーザ属性をユーザの関心事を表すキーワードに変換する属性ポリシーと，システムがサービスとして提供できる知識領域毎にそこで用いられる語の関係を記述した領域オントロジーを持つ．検索時には，ユーザ属性から決まるキーワードを含む文書を選択し，次に，ユーザが指定した領域オントロジーを用いて，ユーザ文書との概念的な類似性が高い順に提示する．市役所の役職と職務内容を記述した属性ポリシー，旅行に関する領域オントロジーを作成し，ユーザ文書としてブログの文書を与えて，朝日新聞1年分を用いて実験を行っている．その結果，意思決定のヒントとなる文書を提示できる可能性を確認している．
キーワード 意思決定支援，文書検索

A decision support system based on document retrieval using user attributes and domain knowledge

Hidefumi OTSUKA† Toshihiro NISHIZONO‡

†‡Graduate School of Engineering, Nihon University
1 Nakagawara, Tokusada, Tamura-machi, Koriyama, Fukushima, 963-842 Japan
E-mail： †otsuka@cscw00.ce.nihon-u.ac.jp ‡nishizono@cs.ce.nihon-u.ac.jp

Abstract This paper proposes a decision support system which retrieves documents containing hints for initial stage of problem solving. At this stage, intentions of users cannot be indicated with accurate keywords to retrieve information necessary for solving problem since alternatives of the solution reach far and wide area. The system provides attribute policies which translate user attributes, such as hobby and official position, into keywords describing user concerns and domain ontology descriptions which specify relationship among terms used in each service knowledge domain. At a retrieve process, the system sorts out retrieved documents containing keywords decided with the user attribute and indicated them in the order of conceptual similarity with the user document calculated through the domain ontology indicated by the user. An experiment is done using an attribute policy of city office, domain ontology in travel knowledge, a blog user document and Asahi-newspaper articles for one year. The experiment ascertains that document retrieval obtaining hints for decision making is possible.
Key words Decision-making support, document retrieval

1. はじめに

情報検索は，目的を達成するための行動を決める際に用いる情報を探す手段である．Web等において広く使われているキーワードを使った検索をする場合，ユーザは明確な意図を持ち，具体的なキーワードを入力することで，検索

システムに意図を伝える．しかし，検索領域の知識が乏しい場合や，問題が抽象的であり解決のための選択肢が広範囲に及ぶ場合（抽象的な意図の場合）には，適確なキーワードが指定できない．

そこで，本稿では，行動の方向性を決めるためのヒントになる文書を示すシステムの提案と検討を行う．

先ず，システムの要求条件を述べ，提案システムで用いる情報とその使い方を示す．次に，ユーザの文書と検索対象文書との類似性の計算手法，提示文書の構造，システムのモジュール構成を示す．最後に実験の結果と考察を示す．

2．システムの要求条件

2．1．意思決定支援

意思決定は，集めた情報を使って，行動を決定するプロセスであるが，その初期段階においては，行動の選択肢が広範囲に及ぶため，必要な情報自体が不明確である．意思決定の初期段階を支援するためには，情報の収集範囲をいくつか例示し，その中から選択させることと，関心事や考えなどのユーザの特性を予め持ち，ユーザの抽象的な意図を推測する必要がある．それらを用いた検索により，行動の方向性を決めるアイディアやヒントになる知識が得られる文書を提示することにより支援する方法が有効だと考えられる．

例えば，市役所の環境政策担当のユーザが，市民の環境保全意識を向上させる政策を決めたい場合を考える．政策には，イベントの企画や罰則の制定など，様々なものが考えられる．まず，その中から方向性を決める必要がある．この様な場合，ユーザが適確なキーワードを指定し，システムに意図を伝えることは困難である．しかし，旅行や映画といった，システムが提示する情報の範囲の中から意図に近いものを選択することはできると考える．そこで，提供できる情報の範囲をシステムが提示し，その中から選択することでシステムに意図を伝える．前述したユーザが範囲として旅行を指定した際に，関心を持っている環境の話題を含んでいる「自然を守って旅をしよう　環境省がエコツアーHP」といった文書を得ることができれば，環境保全意識を向上させる政策のアイディアとして，エコツアーイベントの企画を思いつくと考えられる．政策の方向性が決まることで，より具体的な情報検索と意思決定を行うことができるようになる．

2．2．システムへの要求条件

ユーザは，問題解決の方向性を決めるヒントを得るために本システムを使用する．抽象的な意図のユーザの意思決定を支援するためには，システムが検索領域の知識を持ち，ユーザの関心事や日常の考えを知っている必要がある．

ヒントになるものとして選定される文書は，次の３つの要求条件を満たす必要がある．①ユーザ固有の関心事に関連する．②ユーザの検索したい領域の知識を含む．③日常の考えと関連がある．

①を満たすため，仕事上の役職や趣味など，関心事の知識を表す情報（属性ポリシーと呼ぶ）を準備しておき，ユーザに自身と合致するものを選択させることで，関心事に関連のある文書を選定する．

②を満たすため，旅行や映画といった特定の領域の知識を表す情報（領域オントロジーと呼ぶ）を選択肢として複数準備し，ユーザに検索したい領域を指定させる．即ち，ユーザの抽象的な意図がシステムに伝えられることになる．

③を満たすため，ブログなど，ユーザの日常の考えを表す文書（ユーザ文書と呼ぶ）を持ち，①で選定した文書の中で，②の領域オントロジーの観点から類似性が高いものを選定する．

3．提案システム

図１に，システムで用いる情報の組み合わせと文書提示までの流れを示し，以下に属性ポリシー，領域オントロジー，ユーザ文書と検索対象文書との類似性の計算手法，提

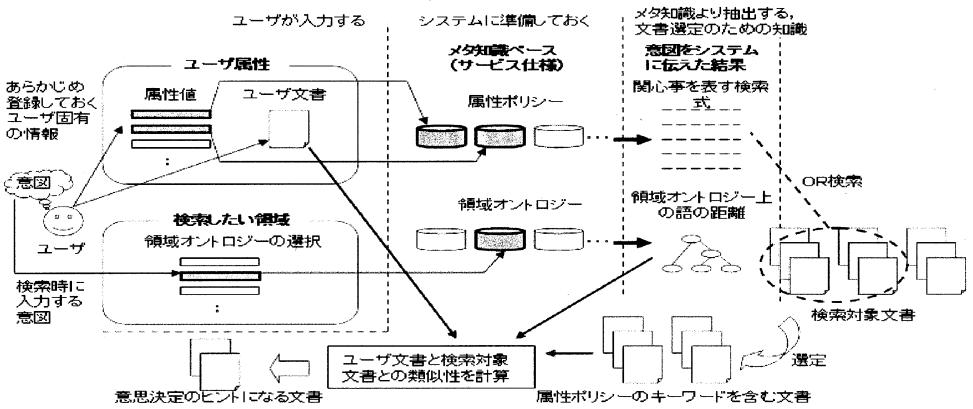

図１．情報の組み合わせと文書提示までの流れ

示文書の構造，システムのモジュール構成を説明する．検索対象文書は，あらかじめ収集され，システムに保存されているとする．キーワード検索や類似性の計算に文書中の名詞を使用するため，日本語形態素解析システム「Chasen」[1]を用いて抽出する．

3．1．属性ポリシー

属性ポリシーは，属性値とそれに対応する検索式を記述したルールの集合である．属性値は，仕事上の役職や趣味などを表すものであり，属性ポリシーの名称と，その中で定義されている条件(役職名など)とする．図1において，ユーザが登録した属性値からユーザ固有の関心事を表す検索式を選ぶために使用する．様々なユーザの条件に対応するため，複数準備する必要がある．

仕事上の役職を属性値とした場合の属性ポリシーの作成手法について説明する．市役所や病院などの組織ごとに作成し，組織内での役職の上下関係を利用する．組織の最下層の役職に，業務内容を示すキーワードを論理積や論理和により組み合わせた検索式を複数与える．それらのいずれかに該当する文書を関心事に関連する文書とする．上位の役職にある人は，下位の役職の検索式を継承する．上位の役職になるほどOR条件で用いる検索式が多くなり，関心事と関係の薄い文書が選ばれる問題がある．そのため，継承した各検索式に役職の役割を表すキーワードを論理積で追加することにより，ノイズとなる文書を除外することを検討している．記述には，アクセス制御言語であるXACML（eXtensible Access Control Markup Language）の使用を検討している．

図2は，Web上で公開されている市役所の組織図と業務内容の説明を参照し，人手で作成した市役所の属性ポリシーの一部である．例えば環境政策担当の場合，その業務内容に環境教育や，環境保全意識などが書いてあるため，それらを組み合わせて図3に示す業務内容を表す検索式を作成した．

3．2．領域オントロジー

旅行，映画，音楽など，特定の領域で使われる語の概念上の上位下位関係を記述したものである．ユーザが指定した領域から，その領域におけるユーザ文書と選定された文書との類似性の強さを計算することに使用する．ユーザが領域オントロジーを指定することで，システムが動作し，文書が提示される．ユーザの様々な意図に対応するために十分な数を準備する必要がある．図4に，旅行についての領域オントロジーの一部を示す．これは旅行をする上で，目的地が重要であるという観点に基づき，シソーラス（日本語語彙大系[2]）を利用して作成したものである．語の関係をSame-as，Is-a，Part-of，で表す．Same-asは，同一であるという「A」＝「B」の関係を表す．Is-aは，「A」は「B」

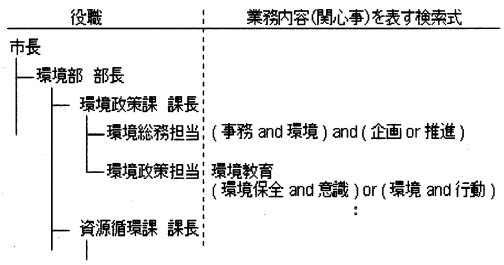

図2．市役所の属性ポリシーの一部

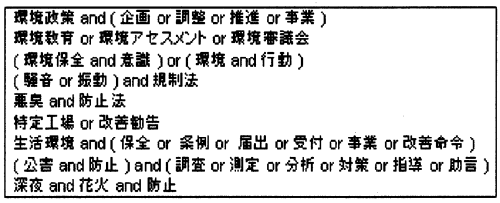

図3．環境政策担当に与えられる検索式

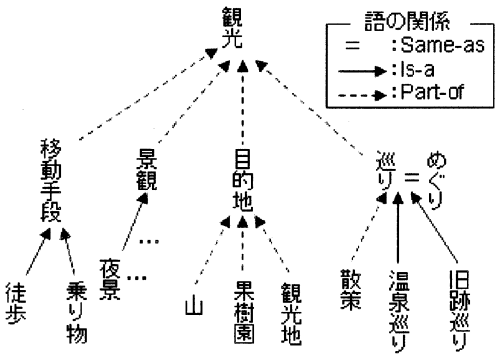

図4．旅行についての領域オントロジーの一部

の一種であるという「A」∈「B」の関係を表す．それ以外の関係はPart-ofで表す．記述にはオントロジー記述言語として用いられるRDF/OWLの使用を検討している．

3．3．類似性の計算手法

ユーザの日常の考えと検索時の意図に関連の強い文書ほど，意思決定のヒントとなる情報が含まれている可能性が高いと考える．そこで，ユーザ文書と属性ポリシーにより選定された文書との領域オントロジー上での類似性を計算する．一般に，二つの文書の類似性を計算する場合，ベクトル空間法の類似度[3]を用いて，文書中の名詞の出現回数を要素とした文書のベクトルを作成し，式(1)によりその内積を計算する方法が採られる．

$$Sim = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}} \qquad (1)$$

x_i：ユーザ文書における名詞iの出現回数

y_i：検索対象文書における名詞iの出現回数

ここでは，領域オントロジー上での類似性を考慮するため，文書中の名詞の出現回数に加え，領域オントロジー上の関連性により出現したと見なされる回数（見なし出現回数と呼ぶ）を要素とする文書のベクトルを作成する．見なし出現回数は，文書中の名詞が領域オントロジー中の語である場合，それに距離が近い語も距離に応じて決まる回数だけ文書に出現したと見なしたものである．見なし出現回数 W_i を式 (2) で定義する．

$$W_i = \sum_j F_j D_{ij} \qquad (2)$$

Fj : 文書中に出現する名詞 j の出現回数

Dij : 名詞 j と領域オントロジー中の語 i との距離

Dij は，領域オントロジー中の語 i と名詞 j との間の領域オントロジー上のパスによって決まるもので，パス上のリンクにおける関係の種類に与えられる係数の積により求める．関係 Same-as，Is-a，Part-of の係数を 1，0.75，0.5 とする．名詞 j が領域オントロジー上に無ければ Dij は 0 となる．なお，距離が遠い語を除くため，Dij が閾値より小さい場合も 0 とする．閾値は，領域オントロジー上における語の距離の平均値 0.1137 を 2 倍した 0.2274 とする．

見なし出現回数を含めた文書のベクトルは，元の文書のベクトルから領域オントロジー中の語を除いたものと，見なし出現回数を要素としたオントロジー中の語のベクトルを結合したものとする．以上より，文書中の名詞 m 個が領域オントロジーの語である場合，文書のベクトル v は以下のようになる．

$$v = (F_1, F_2, \ldots, F_{n-m}, W_1, W_2, \ldots, W_N) \qquad (3)$$

ユーザ文書と検索対象文書との領域オントロジー上の類似性を計算するためには，領域オントロジー中に出現しない文書中の名詞の重みを 0 とする必要がある．このため，領域オントロジー中の語の見なし出現回数を要素とした重みベクトル v'を以下で定義する．

$$v' = (W_1, W_2, \ldots, W_N) \qquad (4)$$

以上を用いて，領域オントロジー上でのユーザ文書と検索対象文書との類似度（オントロジー類似度 OntSim）は次式で与えられる．

$$OntSim = \frac{\sum V'_i v'_i}{\sqrt{\sum V_i^2}\sqrt{\sum v_i^2}} \qquad (5)$$

V : ユーザ文書における語 i の出現回数

v : 検索対象文書における語 i の出現回数

V' : ユーザ文書における語 i の重み

v' : 検索対象文書における語 i の重み

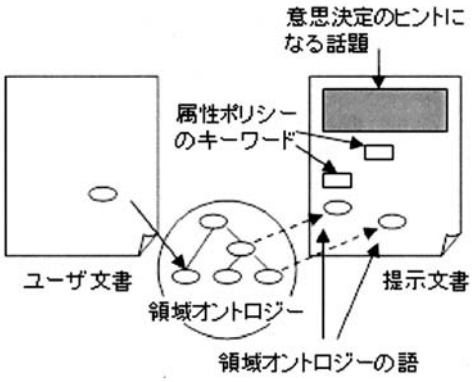

図 5．提示文書の構造

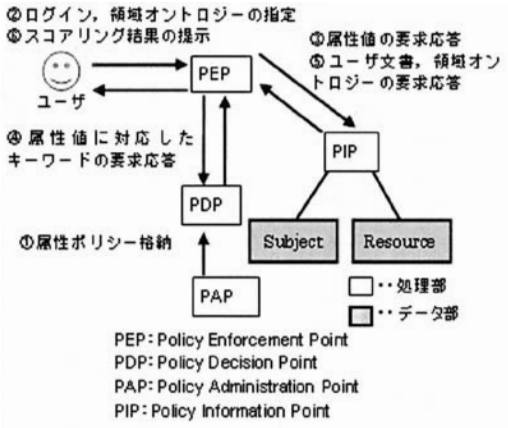

図 6．モジュール構成とデータフロー

式 (5) は，ユーザ文書と検索対象文書に領域オントロジーの語が出現しない場合に値が 0 となり，文書中の名詞が全て領域オントロジーの語に該当する同一文書同士の場合に値が 1 となる．なお，処理の軽減のため，類似性の計算を行う前に領域オントロジーの語を含まない文書はオントロジー類似度が 0 となるため除外する．

3．4．提示文書の構造

オントロジー類似度が上位の文書を意思決定のヒントになる文書として，ユーザに提示する．提示文書の構造を図 5 に示す．提示文書には属性ポリシーのキーワードと領域オントロジーの語が出現する．そのため，ユーザの意図に関連のある話題を含んでいる．また，ユーザの日常の考えの中に現れる領域オントロジーの語との関連性が高い話題も含んでいると考えられる．そのような話題が意思決定のヒントになると考える．

3．5．システムのモジュール構成

XACML のデータフローモデルを利用したシステムのモジュール構成とデータフローを図 6 に示す．データ部には，以下の情報を格納する．

Subject：ユーザ属性（属性値，ユーザ文書）．
Resource：領域オントロジー，検索対象文書．
処理部は，以下の動作を行う．
PEP：ユーザからの入力を受け，属性値を PIP から取得する．その属性値に対応する検索式を PDP に問い合わせる．
PIP から，ユーザ文書，領域オントロジーを取得する．ユーザ文書と類似性が高い文書を提示する．
PDP：属性ポリシーを格納し，指定された属性値から検索式を選ぶ．
PIP：Subject，Resource を格納し，PEP に指定されたユーザ属性や領域オントロジーを送る．指定された属性ポリシーの検索式を使い文書を選定する．
PAP：属性ポリシーを PDP へ設定する．
次にデータフローを説明する．

ユーザ属性を登録する場合，PEP から提供される属性値の選択肢やユーザ文書の入力フォームにより，ユーザが入力した情報は，PIP へ送られ，Subject に格納される．

文書を検索する場合，PEP はログインしたユーザの属性値を PIP から取得し，PDP にその属性値に対応する検索式を問い合わせる．PDP は属性ポリシーから属性値に対応した検索式を選び，それを応答として返す．PEP は検索式に該当する文書を PIP から取得する．次に，PEP はユーザ文書とユーザが指定した領域オントロジーを PIP から取得する．最後に，PEP がそれらを用いてユーザ文書と取得した文書との類似性を計算し，オントロジー類似度の高い順に提示する．

4．実験

提案システムを用いて，ヒントになる文書が提示可能なのかを確認する実験と，考察を述べる．

4．1．実験条件

検索対象文書には，朝日新聞 2004 年[4]の 1 年分（162261 件）を使用する．属性ポリシーには，3．1．節で述べた市役所の役職についての属性ポリシー（最下層の役職が 157 種類あり，平均 5 個の検索式が与えられている）を用いて検索対象文書を選定する．領域オントロジーには 3．2．節で述べた旅行についての領域オントロジー（342 個の語で構成されている）を用いて，選定された検索対象文書に対して，オントロジー類似度を求め，上位 30 件中にヒントになる文書が含まれているかを確認する．また，式 (1) に基づく類似度による順位も求め，順位の違いを比較する．

富山の名水2
富山の名水第2弾です。
3月27日にご紹介したのは、中新川郡上市町の「城山の湧水」でしたが、今日の名水も同じ上市町で穴の谷霊場で湧き出ている「穴の谷(あなんたん)霊水」です。
ここの水は、全国名水100選に選ばれていますので、ご存知の方もいらっしゃるかもしれません。
中略
そういう意味では、健康な家づくりも大事ですが、防腐剤入りの食べ物や飲み物など食生活も見直さないと、本当の健康になれないのかもしれませんね。

図 7．ユーザ文書

実験で使用するユーザの属性と意図を以下に示す．
属性値（関心事）：市役所の環境政策担当．図 3 で示した検索式が適用される．
意図：市民の環境保全意識を向上させる政策を考えたい，とする．
検索したい領域：旅行．
ユーザ文書（日常の考え）：図 7 に示す文書とする．旅行領域オントロジーの語である，料理，車，帰り，名水，を含んでいる．内容から，ユーザが移動に車を使うこと，健康志向であることが分かる．

ヒントになる文書の判断基準を以下に示す．
属性値と，ユーザ文書の内容に関連する話題を含んでおり，その内容から行動の方向性を決めることができる文書とする．上記のユーザの場合，環境と健康に関連する話題を含んでおり，環境意識を高める政策の方向性を決めることができる文書とする．

4．2．結果

属性ポリシーを使った選定により 162261 件から 976 件になった．それらの文書に対して，オントロジー類似度を求め，上位 30 件の文書について，前述した判断条件に基づき，主観でヒントになる文書を確認した．その結果，ヒントになる文書を 3 件発見できた．それらのオントロジー類似度と類似度による順位，タイトル，文書に出現する属性ポリシーのキーワードと領域オントロジーの語を表 1 に示す．

表 1．ヒントになると判断した文書

文書	オントロジー類似度による順位	類似度による順位	タイトル	属性ポリシーのキーワード	領域オントロジーの語
A	1	704	路面電車に復権の動き　排ガスなく、よさ再認識	環境政策　事業　推進	乗り物　タクシー　電車　地下鉄　客
B	13	950	自然守って旅しよう　環境省が「エコツアー」HP	環境教育	旅行　ツアー　観光
C	26	58	化粧品から低公害車まで、尿素は「縁の下の力持ち」	公害　防止	車

4．3．考察

実験結果から，提案システムにより，ヒントになる文書を提示することができる可能性を示すことができたと考えられる．各文書の概要，ヒントになると判断した理由を以下に示す．

文書A

概要:路面電車は排ガスを出さず環境に良い.そのため，路線の新規建設も計画されている．

理由：排ガスについての話題があるため，環境と健康に関連する話題を含むと判断した．また，路面電車が環境に良い，という内容から，環境に良い乗り物を利用して環境保全意識を向上させることを思いつくと考えた．これにより，環境に良い乗り物を導入するなどの方向性を決めることができる．

文書B

概要：自然を傷つけないで，その魅力に触れるエコツアーを集めたホームページを環境省が開設した．

理由：エコツアーに低公害バスを利用する，といった記述があるため，環境と健康に関連する話題を含むと判断した．また，エコツアーが環境教育につながる，という内容から，エコツアーなどのイベントにより環境保全意識を向上させることを思いつくと考えた．これにより，イベントの企画をするといった方向性を決めることができる．

文書C

概要：尿素の用途には，化粧品や排ガスの浄化装置など様々なものがある．

理由：文書Aと同様に，排ガスについての話題があるため，環境と健康に関連する話題を含むと判断した．また，尿素を使った排ガス除去のシステムを搭載したトラックが発売された，という内容から，文書Aと同様に環境に良い乗り物の利用を思いつくと考えた．

以上の文書は，図5に示した構造になっており，文書Aでは路面電車が環境に良いという話題，文書Bではエコツアーが環境教育につながるという話題，文書Cでは排ガス除去システムの話題がユーザの意図に関連すると考えられる．また，文書A，Bは，類似度による順位が低いため，オントロジー類似度でなければ，発見が難しい文書であると言える．即ち，提案システムの有効性を示すことができたと考えられる．

5．まとめ

本稿では，ユーザ属性と検索領域知識を用いて，意思決定のヒントになる文書を提示するシステムを提案し，実験を行った．その結果，数は少ないが，ヒントになる文書を提示できる可能性を示すことができた．しかし，ユーザの意図に全く関係の無い文書が多数含まれていた問題や，実験条件によってはヒントになる文書が発見できない問題がある．前者については，オントロジー類似度の検討を行う必要がある．後者については，属性ポリシーのキーワードや領域オントロジーの語が少ないことが原因と考えられるため，それらの検討と拡充が必要である．

今後の課題として，ヒントになる文書を客観的に判断できる基準の検討を行うこと，また，属性ポリシーや領域オントロジーについて検討と拡充を行うことが挙げられる．

文献

[1]松本祐治 他："日本語形態素解析システム 『茶筌』", NAIST Technical Report，NAIST-IS-97012，1999

[2]池原悟 他："日本語語彙大系 CD－ROM 版" 岩波書店 (1999)

[3]北研二："情報検索アルゴリズム" 共立出版 pp60-61(2002)

[4]朝日新聞社：朝日新聞記事データ集 2004 年版